

CLERK OF THE DISTRICT COURT  
"Express Mail" Mailing Label No. EB 533 1775X  
Date July 24, 1993  
If you are not the addressee, please return this letter to the  
Office of the Clerk of the District Court, 1000 Bank Building,  
San Francisco, California 94102.  
MAURICE L. FUETT  
(Type or Print Name of Person to Whom Letter is Sent)  
[Signature]  
(Signature of Person Mailing Paper or Packet)

5                    IMPROVED EXPRESSION USING FUSED GENES  
                    PROVIDING FOR PROTEIN PRODUCT

CROSS-REFERENCE TO RELATED APPLICATIONS

10                    This application is a continuation-in-part of  
                    U.S. Serial No. 07/680,046, filed 29 March 1991, which is  
                    a continuation of application Serial No. 07/169,833,  
                    filed 17 March 1988, which is a continuation-in-part of  
                    U.S. Serial No. 717,209, filed 28 March 1985, from which  
15                    priority is claimed pursuant to 35 U.S.C. § 120, and  
                    which applications are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Field of the Invention

20                    There are an increasingly large number of genes  
                    available for expression, where the expression product  
                    may find commercial use. In many instances, the initial  
                    expression has been observed in *E. coli*. Expression in  
                    *E. coli* has many disadvantages, one in particular being  
25                    the presence of an enterotoxin which may contaminate the  
                    product and make it unfit to administration to mammals.  
                    Furthermore, there has not previously been an extensive  
                    technology concerned with the production of products in  
                    *E. coli*, as compared to such other microorganisms as  
                    *Bacillus subtilis*, *Streptomyces*, or yeast, such as  
30                    *Saccharomyces*.

                    In many situations, for reasons which have not  
                    been resolved, heterologous products, despite active  
                    promoters and high copy number plasmids, are produced in  
                    only minor amount, if at all, in a microorganism host.  
35                    Since the economics of the processes are dependent upon a

09931216 "081601  
TOP SECRET

substantial proportion of the nutrients being employed in the expression of the desired product, the production of these products in unicellular microorganisms appears to be unpromising. There is, therefore, a substantial need  
5 for processes and systems which greatly enhance the production of a desired polypeptide without substantial detriment to the viability and growth characteristics of the host.

Description of the Prior Art

10 Villa-Komaroff et al., Proc. Natl. Acad. Sci. USA (1978) 75:3727-3731, describes a fusion sequence encoding proinsulin joined to the N-terminus of penicillinase for expression in *E. coli*. Paul et al.,  
15 European J. Cell Biol. (1983) 31:171-174, describe a fusion sequence encoding proinsulin joined to the COOH-terminus or a portion of the tryptophan E gene product for expression in *E. coli*. Goeddel et al., *ibid.* (1979) 76:106-110, describe synthetic genes for human insulin A and B chains fused to *E. coli*  $\beta$ -galactosidase gene to  
20 provide a fused polypeptide in *E. coli*. Stepien et al., Gene (1983) 24:289-297, describe expression of insulin as a fused product in yeast, where the proinsulin gene was fused to the N-terminus coding sequence of *GAL1* for expression in yeast.

25

SUMMARY OF THE INVENTION

Methods and compositions are provided for producing heterologous polypeptides in high yield in a eukaryotic microorganism host, whereby a completely  
30 heterologous fused product is expressed, one part of the peptide being a product shown to be expressed independently in high yield in such host and the remaining part of the product being a polypeptide of interest, resulting in production of the fused product in  
35 high yield. Sequences coding for the two polypeptides

5  
10

## DESCRIPTION OF THE SPECIFIC EMBODIMENTS

15

20

25

30

35

to those polypeptides which are produced in amounts of 5% or greater of the total protein produced by the host, those DNA sequences encoding for such polypeptides may be used in this invention. The DNA sequences may be  
5 identical to the heterologous gene encoding the sequence, may be mutants of the heterologous gene, or may have one or more codons substituted, whereby the codons are selected as being preferred codons by the host. Preferred codons are those codons which are found in  
10 substantially greater than the mathematical probability of finding such codon, based on the degree of degeneracy of the genetic code, in those proteins which are produced in greatest individual abundance in the host. Particularly, in yeast, the glycolytic enzymes may be the  
15 basis for determining the preferred codons.

The entire gene or any portion of the gene may be employed which provides for the desired high yield of polypeptide in the host. Thus, where the stable polypeptide is of lesser economic value than the  
20 polypeptide of interest, it may be desirable to truncate the gene to a fragment which still retains the desirable properties of the entire gene and its polypeptide product, while substantially reducing the proportion of the total fused product which is the stabilizing  
25 polypeptide. As illustrative of a gene encoding a stable polypeptide product in the yeast, is the gene encoding for superoxide dismutase, more particularly human superoxide dismutase, and the gene encoding for ubiquitin.

30 The DNA sequences coding for the two polypeptides, the stabilizing polypeptide and the polypeptide of interest, may be obtained in a variety of ways. The sequences encoding for the polypeptide may be derived from natural sources, where the messenger RNA or  
35 chromosomal DNA may be identified with appropriate

099346 091601

probes, which are complementary to a portion of the coding or non-coding sequence. From messenger RNA, single-stranded (ss) DNA may be prepared employing reverse transcriptase in accordance with conventional techniques. The ss DNA complementary strand may then be used as the template for preparing a second strand to provide double-stranded (ds) cDNA containing the coding region for the polypeptide. Where chromosomal DNA is employed, the region containing the coding region may be detected employing probes, restriction mapped, and by appropriate techniques isolated substantially free of untranslated 5' and 3' regions. Where only portions of the coding sequence are obtained, the remaining portions may be provided by synthesis of adapters which can be ligated to the coding portions and provide for convenient termini for ligation to other sequences providing particular functions or properties.

Where the two genes are obtained in-whole or in-part from naturally occurring sources, it will be necessary to ligate the two genes in proper reading frame. If cleavage of the fused protein is required, where their juncture does not define a selectable cleavage site, genes will be separated by a selectively cleavable site. The selectively cleavable site will depend to some degree on the nature of the genes. That is, the means for cleaving may vary depending upon the amino acid sequence of one or both genes.

Alternatively, there will be situations where cleavage is not necessary and in some situations undesirable. Fused proteins may find use as diagnostic reagents, in affinity columns, as a source for the determination of a sequence, for the production of antibodies using the fused protein as an immunogen, or the like.

The two genes will normally not include introns, since splicing of mRNA is not extensively employed in the eukaryotic unicellular microorganisms of interest.

5           The polypeptide of interest may be any polypeptide, either naturally occurring or synthetic, derived from prokaryotic or eukaryotic sources. Usually, the polypeptide will have at least 15 amino acids (gene of 45 bp), more usually 30 amino acids (gene of 90 bp),  
10       and may be 300 amino acids (gene of 900 bp) or greater.

          Polypeptides of interest include enzymes, fungal, protozoal, bacterial and viral proteins (e.g., proteins from AIDS related virus, such as p18, p25, p31, gp41, etc., and other viral glycoproteins suitable for  
15       use as vaccine antigens), mammalian proteins, such as those involved in regulatory functions, such as lymphokines, cytokines, growth factors, hormones or hormone precursors (e.g., proinsulin, insulin like growth factors, e.g., IGF-I and -II, etc.), etc., blood clotting  
20       factors, clot degrading factors, immunoglobulins, immunomodulators for regulation of the immune response, etc., as well as proteins useful for the production of other biopharmaceuticals.

          The present invention is useful in the  
25       production of viral glycoproteins. For example, the present invention will find use for the expression of a wide variety of proteins from the herpesvirus family, including proteins derived from herpes simplex virus (HSV), varicella zoster virus (VZV), Epstein-Barr virus  
30       (EBV), cytomegalovirus (CMV) and other human herpesviruses such as HHV6 and HHV7. Proteins from other viruses, such as but not limited to, proteins from the hepatitis family of viruses, including hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV),  
35       the delta hepatitis virus (HDV) and hepatitis E virus

099316 091501  
FBI/DOJ

(HEV), as well as retrovirus proteins such as from HTLV-I and HTLV-II and proteins from the human immunodeficiency viruses (HIVs), such as HIV-1 and HIV-2, can also be conveniently expressed using the present system. (See, 5 e.g., Chee et al., Cytomegaloviruses (J.K. McDougall, ed., Springer-Verlag 1990) pp. 125-169, for a review of the protein coding content of cytomegalovirus; McGeoch et al., J. Gen. Virol. (1988) 69:1531-1574, for a discussion of the various HSV-1 encoded proteins; Baer et al., 10 Nature (1984) 310:207-211, for the identification of protein coding sequences in an EBV genome, Davison and Scott, J. Gen. Virol. (1986) 67:1759-1816, for a review of VZV; Houghton et al., Hepatology (1991) 14:381-388, for a discussion of the HCV genome; and Sanchez-Pescador et al., Science (1985) 227:484-492, for an HIV genome.) 15

Fragments or fractions of the polypeptides may be employed where such fragments have physiological activity, e.g., immunological activity such as cross-reactivity with the parent protein, physiological 20 activity as an agonist or antagonist, or the like.

One of the methods for selectable cleavage is cyanogen bromide which is described in U.S. Patent No. 4,366,246. This technique requires the absence of an available methionine other than at the site of cleavage or the ability to selectively distinguish between the 25 methionine to be cleaved and a methionine within the polypeptide sequence. Alternatively, a protease may be employed which recognizes and cleaves at a site identified by a particular type of amino acid. Common 30 proteases include trypsin, chymotrypsin, pepsin, bromelain, papain, or the like. Trypsin is specific for basic amino acids and cleaves on the carboxylic side of the peptide bond for either lysine or arginine. Further, peptidases can be employed which are specific for 35 particular sequences of amino acids, such as those

093116 091601  
093116 091601

peptidases which are involved in the selective cleavage of secretory leader signals from a polypeptide. These enzymes are specific for such sequences which are found with  $\alpha$ -factor and killer toxin in yeast, such as KEX-2  
5 endopeptidase with specificity for pairs of basic residues (Julius et al., Cell (1984) 37:1075-1089). Also, enzymes exist which cleave at specific sequences of amino acids. Bovine enterokinase (Light et al., Anal. Biochem. (1980) 106:199-206) cleaves to the carboxylic  
10 side of lysine or arginine that is preceded by acid residues of aspartic acid, glutamic acid, or carboxymethyl cysteine. Particularly useful is the sequence (Asp)<sub>4</sub> Lys found naturally as part of the activation peptide of trypsinogen in many species. Other  
15 enzymes which recognize and cleave specific sequences include: Collagenase (Germino and Batia, Proc. Natl. Acad. Sci. (1984) 81:4692-4696); factor x (Nagai & Thygersen Nature (1984) 309:810-812); and polyubiquitin processing enzyme (Ozakaynak et al., Nature (1984)  
20 312:663-666), which is also known as ubiquitin-protein hydrolase.

An endogenous yeast ubiquitin processing enzyme accurately cleaves Ub from heterologous fusion proteins containing any of the 20 amino acids at the Ub-protein  
25 junction. The yeast hydrolase responsible for the cleavage of a ubiquitin-heterologous fusion protein has been characterized at the molecular level by cloning and over expression of its gene product. The yeast hydrolase cleaves the junction peptide bond between the C-terminal  
30 Gly<sub>76</sub> of ubiquitin and the heterologous fusion protein rapidly in all cases, except when the first amino acid of the extension protein is proline.

Ubiquitin (Ub), a highly conserved 76 residue protein, is found in eukaryotes either free or covalently  
35 joined via its carboxy-terminal glycine residue to a

099316-091601



variety of cytoplasmic, nuclear and integral membrane proteins. Its use as a stabilizing polypeptide in a heterologous fusion product would allow the production of the polypeptide of interest in the host organism, with  
5 the ability to specifically cleave the ubiquitin-heterologous fusion protein using the Ub-protein hydrolase. Furthermore, the carboxy terminal amino acid sequence of ubiquitin may be incorporated into the fusion product as the cleavable site which links the stabilizing  
10 polypeptide (e.g., SOD) to the polypeptide of interest, thereby affording specific cleavage by Ub-protein hydrolase to liberate the polypeptide of interest.

In addition to the amino acids comprising the cleavable site, it may be advantageous to separate  
15 further the two fused polypeptides. Such a "hinge" would allow for steric flexibility so that the fused polypeptides would be less likely to interfere with each other, thus preventing incorrect folding, blockage of the cleavage site, or the like.

20 The "hinge" amino acid sequence could be of variable length and may contain any amino acid side chains so long as the side chains do not interfere with the mode of action employed to break at the cleavable site or with required interactions in either fused  
25 polypeptide, such as ionic, hydrophobic, or hydrogen bonding. Preferably the amino acids comprising the hinge would have side chains that are neutral and either polar or nonpolar and may include one or more prolines. The hinge region will have at least one amino acid and may  
30 have 20 or more amino acids, usually not more than 15 amino acids, particularly the nonpolar amino acids G, A, P, V, I, L, and the neutral polar amino acids, N, Q, S, and T.

Exemplary hinge sequences may be, but are not  
35 limited to: N-S; Q-A; N-S-G-S-P; A-A-S-T-P; N-S-G-P-T-P-

099416-037.006

P-S-P-G-S-P; S-S-P-G-A; and the like. It is contemplated that such hinge sequences may be employed as repeat units to increase further the separation between the fused polypeptides.

5           So that the "hinge" amino acids are not bound to the final cleaved polypeptide of interest, it is desirable, but not required to practice the invention, to place the "hinge" between the polypeptide that is produced independently at high yield and the sequence for  
10   the cleavable site.

          Where one or more amino acids are involved in the cleavage site, the codons coding for such sequence may be prepared synthetically and ligated to the sequences coding for the polypeptides so as to provide  
15   for a fused protein where all the codons are in the proper reading frame and the selectable cleavage site joins the two polypeptides.

          Instead of only a small portion of the fused coding sequence being synthetically prepared, the entire  
20   sequence may be synthetically prepared. This allows for certain flexibilities in the choice of codons, whereby one can provide for preferred codons, restriction sites, avoid or provide for particular internal structures of the DNA and messenger RNA, and the like.

25           While for the most part, the fused coding sequence will be prepared as a single entity, it should be appreciated that it may be prepared as various fragments, these fragments joined to various untranslated regions, providing for particular functions and  
30   ultimately the coding sequences brought together at a subsequent stage. However, for clarity of presentation, the discussion will be directed primarily to the situation where the coding sequence is prepared as a single entity and then transferred to an expression  
35   vector.

099416-03661  
F09T00-9T000

5 The various sequences comprising the parts of  
the fused coding sequence can be joined by introducing a  
first fragment into a cloning vector. The resulting  
clone may then be restricted at a site internal to the  
coding sequence and an adapter introduced which will  
replace any lost codons and which has a convenient  
terminus for joining to the next fragment. The terminus  
may be cohesive or blunt-ended, depending upon the  
particular nucleotides involved. After cloning of the  
10 combined first fragment and adapter, the vector may be  
restricted at the restriction site provided by the  
adapter and the remaining coding sequence of the second  
fragment introduced into the vector for ligation and  
cloning. The resulting fused sequence should be flanked  
15 by appropriate restriction sites, so that the entire  
sequence may be easily removed from the cloning vector  
for transfer to an expression vector.

The expression vector will be selected so as to  
have an appropriate copy number, as well as providing for  
20 stable extrachromosomal maintenance. Alternatively, the  
vector may contain sequences homologous to the host  
genomic sequences to allow for integration and  
amplification. The expression vector will usually have a  
marker which allows for selection in the expression host.  
25 In order to avoid the use of biocides, which may find use  
in certain situations, desirably, complementation will be  
employed, whereby the host will be an auxotroph and the  
marker will provide for prototrophy. Alternatively, the  
episomal element may provide for a selective advantage,  
30 by providing the host with an enhanced ability to utilize  
an essential nutrient or metabolite in short supply. The  
significant factor is that desirably the extrachromosomal  
cloning vector will provide a selective advantage for the  
host containing the vector as compared to those hosts

which may spontaneously lose the vector during production of the fused polypeptide.

The cloning vector will also include an active transcriptional initiation regulatory region, which does not seriously interfere with the viability of the host. Regions of particular interest will be associated with the expression of enzymes involved in glycolysis; acid phosphatase; heat shock proteins; metallothionein; etc. Enzymes involved with glycolysis include alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, glucose-6-phosphate dehydrogenase, pyruvate kinase, triose phosphate isomerase, phosphofructokinase, etc.

Various transcriptional regulatory regions may be employed involving only the region associated with RNA polymerase binding and transcriptional initiation ("promoter region"), two of such regions in tandem, or a transcriptional initiation regulatory region ("control region"), normally 5'- to the promoter region, where the control region may be normally associated with the promoter or with a different promoter in the wild-type host. The control region will provide for inducible regulation where induction may be as a result of a physical change, e.g., temperature, or chemical change, e.g., change in nutrient or metabolite concentration, such as glucose or tryptophan, or change in pH or in ionic strength.

Of particular interest is the use of hybrid transcriptional initiation regulatory regions. Preferably, the hybrid transcriptional initiation regulatory region will employ a glycolytic enzyme promoter region. The control region may come from the control regions of a variety of expression products of the host, such as ADHII, GAL4, PHO5, or the like.

The transcriptional initiation regulatory  
35 regions may range from about 50-1000 base pairs (bp) of

the region 5' of the wild-type gene. In addition to regions involved with binding of RNA polymerase, other regulatory signals may also be present, such as a capping sequence, transcriptional initiation sequences, enhancer, 5 transcriptional regulatory region for inducible transcription, and the like.

The transcriptional initiation regulatory region will normally be separated from the terminator region by a polylinker, which has a plurality of unique 10 restriction sites, usually at least two, and not more than about 10, usually not more than about six. The polylinker will generally be from about 10-50 bp. The polylinker will be followed by the terminator region, which may be obtained from the same wild-type gene from 15 which the promoter region was obtained or a different wild-type gene, so long as efficient transcription initiation and termination is achieved when the two regions are used.

By digestion of the expression vector with the 20 appropriate restriction enzymes, the polylinker will be cleaved and the open reading frame sequence coding for the fused polypeptide may be inserted. Where the polylinker allows for distinguishable termini, the fused gene can be inserted in a single orientation, while where 25 the termini are the same, insertion of the fused gene will result in plasmids having two different orientations, only one of which will be the proper orientation. In any event, the expression vector may be cloned where it has a prokaryotic replication system for 30 isolation and purification and then introduced into an appropriate eukaryotic host, such as a yeast host. Introduction of foreign DNA into eukaryotic hosts can be performed in a wide variety of ways, such as calcium- 35 polyethylene glycol treated DNA with spheroplasts, use of liposomes, mating, or the like.

The host cells containing the plasmid with the fused gene capable of expression are then grown in an appropriate nutrient medium for the host. Where an inducible transcriptional initiation regulatory region is employed, the host cell may be grown to high density and initiation turned on for expression of the fused polypeptide. Where the promoter is not inducible, then constitutive production of the desired fused polypeptide will occur.

The cells may be grown until there is no further increase in product formation or the ratio of nutrients consumed to product formation falls below a predetermined value, at which time the cells may be harvested, lysed and the fused protein obtained and purified in accordance with conventional techniques. These techniques include chromatography, e.g., HPLC; electrophoresis; extraction; density gradient centrifugation, or the like. Once the fused protein is obtained, it will then be selectively cleaved in accordance with the nature of the selectively cleavable linkage. This has been described previously in relation to the description of the various linkages.

In some instances a secretory leader and processing signal may be included as part of the fused polypeptide. Various secretory leader and processing signals are known, such as yeast  $\alpha$ -factor, yeast killer toxin and the like. The DNA sequence coding for these polypeptide signals may be linked in proper reading frame to the 5'- end (in direction of transcription of the sense strand) of the DNA sequence coding for the fused polypeptide to provide for transcription and translation of a pre-fused polypeptide.

In accordance with the subject invention, the product is produced in at least a 5 weight percent, preferably at least 6 weight percent, and more preferably

at least about 10 weight percent, of the total protein of the host. In this manner, the nutrients employed are efficiently utilized for conversion to a desired product.

The following examples are offered by way of  
5 illustration and not by way of limitation.

## EXPERIMENTAL

EXAMPLE I: Construction and Expression of Expression Vectors for SOD-Proinsulin Fusion Protein

10

### Construction of pYSI1

A yeast expression plasmid pYSI1, containing the human SOD gene fused to the amino-terminus of human proinsulin gene, under the regulation of the GAP promoter and terminator was constructed. A triplet coding for methionine was included between the SOD and proinsulin genes to allow for chemical processing of the fusion protein. The SOD sequences correspond to a cDNA isolated from a human liver library, except for the first 20 codons which were chemically synthesized. The proinsulin sequence was chemically synthesized according to the amino acid sequence reported by (Bell et al., (1979), Nature 282:525-527), but using yeast preferred codons. The GAP promoter and terminator sequences were obtained from the yeast GAP gene (Holland & Holland, J. Biol. Chem. (1979) 254:5466-5474) isolated from a yeast library.

Plasmid pYSI1 was constructed as follows. Three fragments were employed which involve a 454bp *NcoI*-*Sau3A* fragment isolated from phSOD (also designated as pSODNco5), where the fragment includes the entire coding sequence for human superoxide dismutase (hSOD) with the exception of the last three 3'- codons; a 51bp *Sau3A*-*HindIII* synthetic adapter, which codes for the last three codons of hSOD, methionine, and the first 14 codons of

proinsulin; and a 231bp *HindIII*-*SalI* fragment, isolated from pINS5, which encodes proinsulin excepting the first 14 amino acids. These fragments were ligated together and introduced into the plasmid pPGAP, which had been  
5 previously digested with *NcoI* and *SalI* and alkaline phosphatase treated. The resulting plasmid pSI1 was digested with *BamHI* to provide an expression cassette which was cloned into plasmid pC1/1 to yield pYSI1.

Plasmid phSOD (also designated as pSODNco5) is  
10 a pBR322-derived bacterial expression vector which contains a complete cDNA coding (except that the first 20 codons were chemically synthesized) for hSOD as described in copending application Serial Number 609,412 filed on May 11, 1984. Plasmid pINS5 is a pBR322-derived vector  
15 which contains a proinsulin coding sequence chemically synthesized according to the amino acid sequence reported by Bell et al., *Nature* (1979) 282:525-527. Plasmid pPGAP is a pBR322-derived vector described in copending application 609,412 (*supra*) which contains a GAP promoter and GAP terminator (Holland and Holland, *J. Biol. Chem.*  
20 (1979) 254:5466-5474) with a polylinker between them, which provides for single restriction sites for cloning. Plasmid pC1/1 is a yeast expression vector which includes pBR322 sequence, 2 $\mu$  plasmid sequences and the yeast gene  
25 LEU2 as a selectable marker. See EPO 83/306507.1, which relevant parts are incorporated herein by reference.

#### Construction of pYSI2

To prepare the fused gene having the hSOD  
30 coding sequence at the 3'-terminus in the direction of transcription separated from the proinsulin gene by a "spacer" of codons coding for K-R-S-T-S-T-S, the following fragments were ligated. A 671bp *BamHI*-*SalI* fragment containing the GAP promoter, the proinsulin gene  
35 and codons for the spacer amino acids; a 14bp *SalI*-*NcoI*

095416-0101  
TOP SECRET



synthetic adapter, which codes for the last spacer amino acids as a junction of both genes; and a 1.5kb *NcoI*-*BamHI* fragment isolated from pC1/1 GAPSOD described in copending application 609,412 (*supra*), which includes the hSOD coding region, 56bp of hSOD terminator and 934bp of GAP terminator region. The resulting cloned fragment was isolated and inserted into *BamHI* digested, alkaline phosphatase treated pC1/1.

10 Plasmids pPKI1 and pPKI2

Plasmids homologous to pYSI1 and pYSI2, but using the yeast pyruvate kinase (PYK) gene instead of hSOD gene, were also constructed. pPKI1 contains the PYK coding sequence fused to the amino-terminus of the human proinsulin gene under regulation of the yeast PYK promoter and yeast GAP terminator. pPKI2 contains the PYK coding sequence of the 3'-terminus in the direction of transcription separated from the proinsulin gene by a "spacer" of codons coding for K-R-S-T-S. This fused gene is under regulation of the GAP promoter and PYK terminator.

Construction of pYASI1

This yeast expression plasmid is similar to pYSI1 and contains the hSOD gene fused to the amino terminus of the human proinsulin gene, with a methionine codon at the junction between both genes. The fusion gene is under control of the hybrid inducible ADH2-GAP (yeast alcohol dehydrogenase 2) promoter and the GAP terminator. An about 3kbp *BamHI* expression cassette was constructed by replacing the GAP promoter sequence from pYSI1 with the hybrid ADH2-GAP promoter sequence.

The ADH2 portion of the promoter was constructed by cutting a plasmid containing the wild type ADH2 gene (plasmid pADR2, see Beier and Young, Nature

(1982) 300:724-728) with the restriction enzyme *EcoR5*, which cuts at a position +66 relative to the ATG start codon, as well as in two other sites in pADR2, outside of the ADH2 region. The resulting mixture of a vector  
5 fragment and two smaller fragments was resected with *Bal31* exonuclease to remove about 300bp. Synthetic *XhoI* linkers were ligated onto the *Bal31* treated DNA. The resulting DNA linker vector fragment (about 5kb) was separated from the linkers by column chromatography, cut  
10 with the restriction enzyme *XhoI*, religated and used to transform *E. coli* to ampicillin resistance. The positions of the *XhoI* linker additions were determined by DNA sequencing. One plasmid which contained an *XhoI* linker located within the 5' non-transcribed region of  
15 the ADH2 gene (position -232 from ATG) was cut with the restriction enzyme *XhoI*, treated with nuclease *S1*, and subsequently treated with the restriction enzyme *EcoRI* to create a linear vector molecule having one blunt end at the site of the *XhoI* linker and an *EcoRI* end.

20 The GAP portion of the promoter was constructed by cutting plasmid pPGAP (*supra*) with the enzymes *BamHI* and *EcoRI*. followed by the isolation of the 0.4Kbp DNA fragment. The purified fragment was cut with the enzyme *AluI* to create a blunt end near the *BamHI* site.

25 Plasmid pJS104 was constructed by the ligation of the *AluI-EcoRI* GAP promoter fragment to the ADH2 fragment present on the linear vector described above.

30 Plasmid pJS104 was digested with *BamHI* (which cuts upstream of the ADH2 region) and with *EcoI* (which cuts downstream of the GAP region). The about 1.3Kbp fragment containing the ADH2-GAP promoter was gel purified and ligated to an about 1.7Kbp fragment containing the hSOD-proinsulin fusion DNA sequences and GAP terminator present in pYSI1 (previously described).

35

This 3Kbp expression cassette was cloned into *Bam*HI digested and phosphatase treated pC1/1 to yield-pYASI1.

Construction of pYASI1 Derivatives Containing Trypsin and Enterokinase Cleavage Sites

5 A series of plasmids were constructed derived from PYASI1, in which the GAP terminator was replaced by the  $\alpha$ -factor terminator (Brake et al., Proc. Natl. Acad. Sci. USA (1984) 81:4642) and the cleavage site between  
10 SOD and proinsulin was modified to code for trypsin or enterokinase processing sites. Sequences coding for Lys-Arg were used to replace the methionine codon in pYASI1 yielding a trypsin site. Alternatively, sequences coding for (Asp)<sub>4</sub>Lys were used at the cleavage site to yield an  
15 enterokinase site. In addition, sequences coding for extra hinge amino acids were also inserted between the SOD and the cleavage site in other constructions.

Expression of Fusion Proteins

20 Yeast strain 2150-2-3 (*Mat a*, *ade 1*, *leu 2-04*, *cir*<sup>o</sup>) or P017 (*Mat a*, *leu 2-04*, *cir*<sup>o</sup>) were transformed with the different vectors according to Hinnen et al., Proc. Natl. Acad. Sci. USA (1978) 75:1929-1933. Single transformant colonies harboring constitutive GAP  
25 regulated vectors were grown in 2 ml of *leu*<sup>-</sup> selective media to late log or stationary phase. Cells harboring inducible ADH2-GAP regulated vectors were grown to saturation in *leu*<sup>-</sup> selective media, subsequently diluted 1:20 (v/v) in YEP, 3% ethanol, with or without 2 - 3.5mM  
30 CuSO<sub>4</sub> and grown to saturation in this medium. Cells were lysed in the presence of SDS and reducing agent and the lysates clarified by centrifugation. Cleared lysates were subjected to polyacrylamide gel electrophoreses (Laemmli, Nature (1970) 277:680). Following staining  
35 with Coomassie blue, a band of about 28kDal (kilodaltons)

was observed, the size predicted for the fusion protein. This band was detected in those cells transformed with expression vectors, while being absent from extracts of cells harboring control (pC1/1) plasmids. Amount of protein per band was determined by densitometric analysis of the Coomassie blue stained gels. The fusion protein accounts for over 10% of the total cell protein as estimated from the stained gels in those cells transformed with pYSI1, pYSI2 or pYASI1, while it accounts for less than 0.5% in pYPKI1 or pYPKI2 transformants (See Table 1).

TABLE 1

The Yield of SOD-PI from 2150 or P017 Transformed with  
Different Expression Plasmids and Grown in the  
Absence/Presence of 2 - 3.5 mM CuSO<sub>4</sub>.

| Strain                                   | Plasmid | Description of sequences<br>contained in the expression                                  | Expression (percent of<br>total cell protein) <sub>1</sub><br>-Cu <sup>++</sup><br>+Cu <sup>++</sup> |
|--|---------|--|--|
| 2150                                     | pYPKI 1 | PYK <sub>p</sub> PYK M BCA5 GAP <sub>t</sub>   | 0.5  |
| 2150                                     | pYPKI 2 | GAP <sub>p</sub> M BCA5 KRSTS <sub>2</sub> PYK PYK <sub>t</sub>                          | 0.5  |
| 2150                                     | pYSI 1  | GAP <sub>p</sub> SOD M BCA5 GAP <sub>t</sub>   | 10   |
| 2150                                     | pYSI 2  | GAP <sub>p</sub> M BCA5 KR(ST) <sub>2</sub> S SOD GAP <sub>t</sub>                       | 10   |
| <u>500 Met Proinsulin</u>                |         |  |  |
| 2150                                     | pYASI 1 | (ADH-GAP) <sub>p</sub> SOD M BCA5 GAP <sub>t</sub>                                       | 10   |
| P017                                     | pYASI 1 | (ADH-GAP) <sub>p</sub> SOD M BCA5 GAP <sub>t</sub>                                       | 20-30  |
| <u>SOD(hinge) (Asp)<sub>4</sub>LysPI</u> |         |  |  |
| P017                                     | pYSI12  | (ADH-GAP) <sub>p</sub> SOD-D <sub>4</sub> K-BCA5 α-factor <sub>t</sub>                   | 6-9  |
| P017                                     | pYSI15  | (ADH-GAP) <sub>p</sub> SOD-(NS)D <sub>4</sub> K-BCA5<br>α-factor <sub>t</sub>            | 5-6  |
| P017                                     | pYSI8   | (ADH-GAP) <sub>p</sub> SOD-(NSGSP)D <sub>4</sub> K-BCA5<br>α-factor <sub>t</sub>         | 5  |
| P017                                     | pYSI4   | (ADH-GAP) <sub>p</sub> SOD-(NSGTPPPSPGSP)D <sub>4</sub> K-<br>BCAS α-factor <sub>t</sub> | 9-12   |

35 30 25 20 15 10 5

TABLE 1 (continued)

| Strain | Plasmid | Description of sequences contained in the expression                   | Expression (percent of total cell protein) <sup>1</sup><br>-Cu <sup>++</sup><br>+Cu <sup>++</sup> |
|--------|---------|--|---|
|        |         | <u>SOD(hinge) LysArgPI</u>   |   |
| P017   | pYSI13  | (ADH-GAP) <sub>p</sub> SOD-KR-BCA5 α-factor <sub>t</sub>               | 8-10 8-10   |
| P017   | pYSI10  | (ADH-GAP) <sub>p</sub> SOD-(NSGSP)KR-BCA5 α-factor <sub>t</sub>        | 5-7 10-15   |
| P017   | pYSI3   | (ADH-GAP) <sub>p</sub> SOD-(NSGTPPPSPGSP)KR-BCA5 α-factor <sub>t</sub> | 5-8 15-30   |

<sup>1</sup> Determined by scanning densitometer analysis of Coomassie Blue stained gels.

Note: Proinsulin (PI) accounts for less than 0.1% of total cell protein in cells transformed with pYGAPINS5, a plasmid containing the proinsulin gene under regulation of GAPDH promoter and terminator (GAP<sub>p</sub> M BCA5 GAP<sub>t</sub>).

PYK : pyruvate kinase gene GAP<sub>p</sub> : GAP promoter PYK<sub>t</sub> : PYK terminator  
SOD : human SOD gene GAP<sub>t</sub> : GAP terminator (ADH2-GAP)<sub>p</sub> : hybrid  
BCA5 : proinsulin gene PYK<sub>p</sub> : PYK promoter ADH2-GAP promoter  
P, G, D, N, M, K, R, S, T: one letter α-factor<sub>t</sub> : α-factor terminator  
amino acid code

Results shown in Table 1 indicate that while expression levels of PYK-proinsulin fusion are comparable to those obtained with proinsulin alone (about 0.5% and 0.1%, respectively), the expression levels of hSOD-  
5 proinsulin are about 20 to 100 fold higher. The inducible ADH2-GAP hybrid transcriptional initiation regulatory region is preferred, since it is noted that constitutive production in scaled-up cultures results in unstable expression.

10 The hSOD-proinsulin proteins synthesized by yeast were also submitted to Western analysis. Cleared yeast lysates prepared as described above were electrophoresed on polyacrylamide gels (Laemmli, *supra*) and proteins were subsequently electroblotted onto  
15 nitrocellulose filters (Towbin et al., Proc. Natl. Acad. Sci. USA (1979) 76:3450). Two identical filters were blotted. The filters were preincubated for one hour with 1% BSA in PBS and subsequently treated with rabbit anti-hSOD or guinea pig anti-insulin antibodies for 12 hours  
20 at 4°C. Both sera had been preadsorped with pC1/1 control lysate in 10% goat serum. The filters were washed with 1% BSA PBS and a second goat anti-rabbit or anti-guinea pig antibody conjugated with horseradish peroxidase added. Finally, the filters were incubated  
25 with horseradish peroxidase color development reagent (Bio-Rad) and washed. The Western analysis showed that the fusion protein reacted with both antibodies.

#### Cleavage of the Fusion Proteins

30 A saturated culture of 2150 (pYASII1) was grown in SDC minus leucine plus threonine and adenine, containing 2% glucose. This was used to inoculate a 10 liter fermentor containing YEP with 3% ethanol as carbon source. After 48 hours at 30°C, the cells were harvested

by centrifugation (Sharples), weighed (124g), and washed with cold water.

The cells were lysed by glass bead disruption (Dyno mill) using a buffer containing 10mM Tris Cl, pH 7.0, 1mM EDTA, 1μg/ml pepstatin A and 1mM PMSF. The mixture was centrifuged for 20 minutes at 8,000 rpm in a JA10 rotor (Beckman). The pellet was resuspended in 100mls of buffer and the liquid was removed from the beads. This was repeated until ~500mls of buffer was used to thoroughly remove all pellet material from the glass beads. The resuspended pellet was centrifuged, and the pellet washed a second time. The pellet was then extracted for 30 minutes in buffer plus 1% SDS.

The SDS soluble fraction was ion-pair extracted using 500mls of solvent A (Konigsberg and Henderson, (1983) Meth. in Enz. 91:254-259), the pellet washed once with solvent A, and once with acetone.

After drying in a vacuum desiccator, the powder was dissolved in 140mls 100% formic acid. Sixty mls of H<sub>2</sub>O and 20g CNBr were added. After 24 hours at room temperature, in the dark, an additional 20g CNBr was added, and the reaction continued for 24 hours. At this time, the material was dialyzed overnight against 4 liters H<sub>2</sub>O using 2000 MW cutoff tubing (Spectrapor). A second dialysis against 0.1% acetic acid followed. After lyophilization, a powder consisting mostly of SOD-homoserine lactone and proinsulin was obtained, weighing 1.1g.

This powder was dissolved in a 200ml solution of 7% urea, 9% sodium sulfite, and 8.1% sodium tetrathionate - 2H<sub>2</sub>O, pH 7.5. After incubation for 3 hours at 37°C, the S-sulfonate products were dialyzed twice versus 10mM Tris pH 8.0, and once versus 20mM TEAB (triethylammonium bicarbonate) pH 7.3.



5 The S-sulfonates were recovered by lyophilization and dissolved in 240mls DEAE column buffer (Wetzel et al., Gene (1981) 17:63-71) and loaded onto a 60ml column. After washing with two column volumes, the proinsulin-S-sulfonate was eluted with a 600ml gradient of 0 to 0.4M NaCl in column buffer. Fractions containing proinsulin-S-sulfonate were pooled and dialyzed twice against 10mM Tris, pH 7.5, and once against 1mM Tris.

10 The product, ~90% pure proinsulin-S-sulfonate, was shown to migrate as expected on pH 9 gel electrophoresis (Linde et al., Anal. Biochem. (1980) 107:165-176), and has the correct 15 N-terminal residues. On analysis, the amino acid composition was very close to that expected, not exactly correct due to the presence of  
15 a low level of impurities. The yield was 150mg.

Preliminary results on renaturation have been obtained with the following procedure. The proinsulin-S-sulfonate can be renatured at pH 10.5, with  $\beta$ -mercaptoethanol (Frank et al., (1981) in Peptides: Synthesis, Structure and Functions, Proc. of the Seventh Amer. Pep. Symposium, Rich and Gross, eds., Pierce Chemical Co., Rockford, IL, pp.729-738). In preliminary experiments, the yield of correctly renatured proinsulin has been monitored by the production of insulin produced  
20 from digestion with trypsin and carboxypeptidase B. The proinsulin - S - SO<sub>3</sub> produced by this process appears to renature as well as purified porcine proinsulin - S - SO<sub>3</sub>. This process has been reported to yield 70% of the expected amount of insulin. The insulin produced in this  
25 way has the correct N-terminal 15 residues of each A chain and B chain as determined by amino acid sequencing.  
30

EXAMPLE II: Construction and Expression of Expression  
Vectors for SOD-p31 Fusion Protein

A yeast expression plasmid pC1/1-pSP31-GAP-  
ADH2, containing the human SOD gene fused to the amino  
terminus of the endonuclease region (p31) of the *pol* gene  
of the AIDS related virus (ARV) (Sanchez-Pescador et al.,  
Science (1985) 227:484) was constructed. Expression of  
SOD-p31 is non-constitutive and is under regulation of a  
hybrid ADH-GAP promoter.

Construction of pC1/1-pSP31-GAP-ADH2 Derivative

For the construction of a gene for a fused  
protein SOD-p31 to be expressed in yeast, a plasmid  
(pS14/39-2) was used. This plasmid contains the SOD gene  
fused to the proinsulin gene under the regulation of the  
ADH-2/GAP promoter in the same manner as pYAS1. The  
proinsulin gene is located between *EcoRI* and *SalI*  
restriction sites. To substitute the proinsulin gene  
with the p31 fragment, two oligomers designated ARB-300  
and ARV-301, respectively, were synthesized using  
phosphoramidite chemistry. The sequences generate  
cohesive ends for *EcoRI* and *NcoI* on each side of the  
molecule when the two oligomers are annealed. ARV-300  
and ARV-301 have the sequences:

ARB-300 5' AATTCAGGTGTTGGAGC

GTCCACAACCTCGGTAC 5' ARV-301

Two  $\mu$ g of pS14/39-2 linearized with *EcoRI* were  
ligated to 100 picomoles each of phosphorylated ARV-300  
and dephosphorylated ARV-301 in the presence of ATP and  
T4 DNA ligase in a final volume of 35  $\mu$ l. The reaction  
was carried out at 14°C for 18 hours. The DNA was  
further digested with *SalI* and the fragments were  
resolved on a 1% low melting point agarose gel and a  
fragment containing the vector plus the SOD gene (~6.5kb)  
was purified as described above and resuspended in 50  $\mu$ l

of TE (10mM Tris, 1mM EDTA, pH 8). Five  $\mu$ l of this preparation were ligated to 5  $\mu$ l of the p31 fragment (ARV248NL, see below) in 20  $\mu$ l final volume for 18 hours at 14°C and 5  $\mu$ l used to transform competent HB101 cells.

5 The resultant plasmid was called pSP31. Twenty  $\mu$ g of this plasmid were digested with *Bam*HI and a fragment of about 2900 bp was isolated by gel electrophoresis, resuspended in TE and ligated to pC1/1 previously cut with *Bam*HI. This DNA was used to transform HB101 and

10 transformants with the *Bam*HI cassette were obtained. Yeast strain P017 (Mat a, leu2-04, cir<sup>o</sup>) was transformed with this pC1/1-pSP31-GAP-ADH2 derivative.

Preparation of ARV248NL, the p31 Coding Fragment.

15 The 800bp ARV248NL fragment codes for numbered amino acids 737 to the end of the pol protein as shown in Figure 2 of Sanchez-Pescador et al. (*supra*). The following procedure was used for its preparation.

A 5.2kb DNA fragment was isolated from a *Kpn*I

20 digest of ARV-2 (9B) (Sanchez-Pescador et al., *supra*) containing the 3' end of the pol gene, orf-1, env and the 5' end of orf-2, that had been run on a 1% low melting point agarose (Sea-Pack) gel and extracted with phenol at 65°C, precipitated with 100% ethanol and resuspended in

25 TE. Eight  $\mu$ l of this material were further digested with *Sst*I for 1 hour at 37°C in a final volume of 10  $\mu$ l. After heat inactivation of the enzyme, 1.25  $\mu$ l of this digest were ligated to 20 ng of M13mp19 previously cut with *Kpn*I and *Sst*I, in the presence of ATP and in a final

30 volume of 20  $\mu$ l. The reaction was allowed to proceed for 2 hours at room temperature. Five  $\mu$ l of this mixture were used to transform competent *E. coli* JM101. Clear plaques were grown and single-stranded DNA was prepared as described in Messing and Vieira, Gene (1982) 19:269-

35 276.

5 The DNA sequence in the M13 template was  
altered by site specific mutagenesis to generate a  
restriction site recognized by *Nco*I (CCATGG). An  
oligodeoxynucleotide that substitutes the A for a C at  
10 position 3845 (Figure 1 in Sanchez-Pescador *et al.*,  
*supra*) and changes a T for an A at position 3851 was  
synthesized using solid phase phosphoramidite chemistry.  
Both of these changes are silent in terms of the amino  
acid sequences, and the second one was introduced to  
15 decrease the stability of the heterologous molecules.  
The oligomer was named ARV-216 and has the sequence  
5'-TTAAAATCACTTGCCATGGCTCTCCAATTACTG  
and corresponds to the noncoding strand since the M13  
derivative template 01100484 is single-stranded and  
20 contains the coding strand. The 5' dephosphorylated M13  
sequencing primer, 50 mM Tris-HCl pH 8, 20 mM KCl, 7 mM  
MgCl<sub>2</sub> and 0.1 mM EDTA. The polymerization reaction was  
done in 100  $\mu$ l containing 50 ng/ $\mu$ l DNA duplex, 150  $\mu$ M  
dNTPs, 1 mM ATP, 33 mM Tris-acetate pH 7.8, 66 mM  
25 potassium acetate, 10 mM magnesium acetate, 5 mM  
dithiothreitol (DTT), 12.5 units of T4 polymerase, 100  
 $\mu$ g/ml T4 gene 32 protein and 5 units of T4 DNA ligase.  
The reaction was incubated at 30°C for 30 minutes and was  
stopped by the addition of EDTA and SDS (10mM and 0.2%  
30 respectively, final concentration). Competent JM101 *E.*  
*coli* cells were transformed with 1, 2, and 4  $\mu$ l of a 1:10  
dilution of the polymerization product and plated into YT  
plates. Plaques were lifted by adsorption to  
nitrocellulose filters and denatured in 0.2 N NaOH, 1.5 M  
35 NaCl, followed by neutralization in 0.5 M Tris-HCl pH  
7.3, 3 M NaCl and equilibrated in 6 x SSC. The filters  
were blotted dry, baked at 80°C for 2 hours and  
preannealed at 37°C in 0.2% SDS, 10 x Denhardt's 6 x SSC.  
After 1 hour, 7.5 x 10<sup>6</sup> cpm of labelled ARV-216 were  
added to the filters and incubated for 2 additional hours

Replicative form (RF) of the M13 01031098 template was prepared by growing 6 clear plaques, each in 1.5 ml of 2 x YT (0.5% yeast extract, 0.8% tryptone, 0.5% NaCl, 1.5% agar) at 37°C for 5 hours. Double-stranded DNA was obtained as described by Maniatis, et al., Molecular Cloning, a Laboratory Manual, Cold Spring Harbor, (1982) pooled and resuspended in 100 µl final volume. A 20 µl aliquot of RF was cut with *Nco*I and *Sal*I in a 40 µl volume of digestion buffer. This fragment was used for p31 expression in yeast. The samples were run on a 1% low melting point agarose (Sea-Pack) gel and the DNAs were visualized by fluorescence with ethidium bromide. The 800 bp band was cut and the DNA was

extracted from the gel as mentioned above and resuspended in 10  $\mu$ l of TE. The fragment was called ARV248NL.

Induction of pC1/1-pSP31-GAP-ADH2

5 Three different kinds of inductions were tried:

1) P017 colonies were induced in either a 10 ml culture of YEP/1% glucose or a leu<sup>-</sup>/3% ethanol culture for 24 hours. The yeast pellets from each mixture were analyzed for p31 by both polyacrylamide gels and Westerns  
10 using sera from AIDS patients. Even though the Coomassie-stained gel showed a negative result, in both cases the Western did light up a band of the correct molecular weight.

2) P017 colonies were induced in a 30 ml  
15 culture of YEP/1% ethanol for 48 hours. Aliquots were analyzed by PAGE at various time points during the induction. The Coomassie-stained gel shows a band in the correct molecular weight range (47-50 kd) that appears after 14 hours in YEP/1% ethanol and reaches a maximum  
20 intensity at 24 hours of induction. The Western result for SOD p31 using sera from AIDS patients correlates well with the Coomassie-stained gel, showing strong bands at 24 and 48 hours of induction.

25 Purification and Characterization of SOD-p31 from Yeast

Frozen yeast (bacteria) cells were thawed at room temperature and suspended in 1.5 volumes of lysis buffer (20 mM Tris-Cl, pH 8.0, 2 mM EDTA, 1 mM phenylmethylsulfonyl fluoride (PMSF), for bacteria; 50 mM  
30 Tris-Cl, pH 8.0, 2 mM EDTA, 1 mM PMSF for yeast), and mixed with 1 volume of acid-washed glass beads.

Cells were broken for 15 minutes in a non-continuous mode using the glass chamber of a Dynomill unit at 3,000 rpm, connected to a -20°C cooling unit.  
35 Glass beads were decanted for 2-3 minutes on ice, and the

2025 RELEASE UNDER E.O. 14176

cell lysate removed. The decanted glass beads were washed twice with 30 ml of lysis buffer at 4°C. The cell lysate was centrifuged at 39,000 x g for 30 minutes.

The pellet obtained from the above  
5 centrifugation was washed once with lysis buffer, after vortexing and suspending it at 4°C (same centrifugation as above). The washed pellet was treated with 0.2% SDS (for bacteria) and 0.1% SDS (for yeast) in lysis buffer and was agitated by rocking at 4°C for 10 minutes. The  
10 lysate was centrifuged at 39,000 x g for 30 minutes. The pellet was boiled in sample buffer (67.5 mM Tris-Cl, pH 7.0, 5%  $\beta$ -mercaptoethanol, 2.3% SDS) for 10 minutes and centrifuged for 10 minutes at 39,000 x g. The supernatant was recovered and passed through a 0.45  $\mu$ m  
15 filter. The supernatant from the above filter was loaded (maximum 50 mg of protein) on a gel filtration column (2.5 x 90 cm, ADA 34 LKB) with a flow rate of 0.3 - 0.4 ml/min, equilibrated with phosphate-buffered saline (PBS), 0.1% SDS. The fractions containing SOD-p31 were  
20 pooled and concentrated either by vacuum dialysis or using a YM5 Amicon membrane at 40 psi. The protein was stored at -20°C as concentrated solution.

Gel electrophoresis analysis showed that the  
25 SOD-p31 protein migrates having a molecular weight of about 46 kd and is over 90% pure.

Similar constructions and results have been obtained by expressing an SOD-p31 fusion under regulation of a bacterial tap-la promoter in *E. coli*.

The SOD-p31 fused protein finds use in  
30 immunoassays to detect the presence of antibodies against AIDS in body fluids. Successful results have been obtained using the SOD-p31 fusion protein in ELISA as well as in strip assays.

EXAMPLE III: Construction and Expression of Expression  
Vectors for SOD-IGF-2 Fusion Protein.

A yeast expression plasmid pYLUIGF2-14,  
containing the human SOD gene fused to the amino terminus  
5 of the IGF2 gene (see EPO 123 228) was constructed.  
Expression of SOD-IGF2 is non-constitutive and it is  
under regulation of a hybrid ADH-GAP promoter.

Construction of pYLUIGF2-14

10 For the construction of a gene for a fused  
protein SOD-IGF2 to be expressed in yeast, plasmid pYS18  
was used. Plasmid pYS18 contains the SOD gene fused to  
the proinsulin gene under the regulation of the ADH-GAP  
promoter and  $\alpha$ -factor terminator (see Table 1). Plasmid  
15 pYS18 was digested with *Bam*HI and *Eco*RI. The 1830 bp  
fragment (containing the ADH-GAP promoter and SOD gene)  
was purified by gel electrophoresis.

A second *Bam*HI (460 bp) fragment coding for  
amino acid residue 41 to 201 of IGF-2 and for the  $\alpha$ -  
20 factor terminator (see EPO 123 228) was ligated to the  
following linker:

*Eco*RI *Sal*I  
AATTCCATGGCTTACAGACCATCCGAAACCTTGTGTGGTGGTGAATTGG

GGTACCGAATGTCTGGTAGGCTTTGGAACACACCACCACTTAACCAGCT

25 The linker provides for an *Eco*RI overhang, an ATG codon  
for methionine and for codons 1-40 of IGF2 and *Sal*I  
overhang.

The resulting *Eco*RI-*Bam*HI (510 bp) fragment  
containing the IGF-2 gene and  $\alpha$ -factor terminator was  
30 ligated to the 1830 bp *Bam*HI-*Eco*-RI fragment containing  
the ADH-GAP promoter and SOD (see above). The resulting  
*Bam*HI (2340 bp) fragment was cloned into *Bam*HI digested  
and phosphatase treated pAB24 (see below) to yield  
pYLUIGF2-14.

099116-0001  
PAGE 0001



It also contains the yeast URA3 gene derived from plasmid YEp24 (Botstein et al., (1979) Gene 8:17) and the yeast LEU2<sup>d</sup> gene derived from plasmid pC1/1 (see EPO 116201). Insertion of the expression cassette was in the *Bam*HI site of pBR322, thus interrupting the gene for bacterial resistance to tetracycline.

15  
20  
25  
30

35

Two other constructions: pYLUIGF2-15 and pYUIGF2-13 were used as controls for expression of a non-fused IGF2. The former plasmid (pYLUIGF2-15) for intracellular expression contains the IGF2 gene under control of the GAP promoter and  $\alpha$ -factor terminator. The latter plasmid (pYUIGF2-13) for secretion of IGF2, and the IGF-2 gene under control for the GAP promoter,  $\alpha$ -factor leader and  $\alpha$ -factor terminator.

35 30 25 20 15 10 5

TABLE 2

EXPRESSION OF IGF2

| Construction in AB110   | PAGE STAIN<br>% of total cell protein | RRA <sup>+</sup> |
|---|---------------------------------------|------------------|
| 1. pYLUIGF2-15<br>(GAP <sub>p</sub> IGF2·αF <sub>t</sub> )                  | NOT DETECTABLE                        | NA               |
| 2. pYLUIGF2-13<br>(GAP <sub>p</sub> αF <sub>L</sub> ·IGF2·αF <sub>T</sub> ) | BARELY DETECTABLE                     | 10μg/l           |
| 3. pYLUIGF2-14<br>(ADH2/GAP <sub>p</sub> SOD·IGF2·αF <sub>T</sub> )         | 10-15%                                | NA*              |

NA: Not available.

\* By Coomassie blue staining, the SOD·IGF2 fusion protein represents 10-15% of the total cell protein, i.e., - 100-300 mg/l culture equivalent. IGF2 represents -1/3 of the fusion protein, therefore it constitutes about 30-100 mg/l culture equivalent. Analytical CNBr cleavage reactions with the fusion protein have resulted in a band on PAGE which migrates to the position expected for IGF2.

+ RRA-IGF2 levels were measured by a placental membrane radioreceptor assay (RRA) according to Horner et al., J. of Clinical Endocrinology and Metabolism (1978) 47:1287 and Marshall et al., J. of Clinical Endocrinology and Metabolism (1974) 39:283. Placental membranes for the RRA were prepared by the method of Cuatrecasas, Proc. Natl. Acad. Sci. USA (1972) 69:318

Protocol for CNBr Cleavage of SOD-IGF2

The insoluble fraction from glass bead lysis of yeast cells was dissolved in 70% formic acid. CNBr crystals (~g CNBr/100 mg fusion protein) were added and  
5 incubation was carried out at room temperature for 12 - 15 hours in the dark. This step may be repeated after 24 hours if cleavage is incomplete.

It is evident from the above results that otherwise difficultly and inefficiently produced  
10 polypeptides may be produced in substantially enhanced yields by employing a fused protein, where the fusion protein includes a relatively short stable polypeptide sequence joined to the other polypeptide by a selectively cleavable site. Thus, high levels of the fusion protein  
15 are obtained in a eukaryotic host, such as yeast, allowing for the efficient production of desired polypeptides heterologous to the host.

Although the foregoing invention has been described in some detail by way of illustration and  
20 example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

Deposits of Strains Useful in Practicing the Invention

25 A deposit of biologically pure cultures of the following strains was made with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland. The accession number indicated was assigned after successful viability testing, and the requisite fees were  
30 paid. Access to said cultures will be available during pendency of the patent application to one determined by the Commissioner to be entitled thereto under 37 CFR 1.14 and 35 USC 122. All restriction on availability of said cultures to the public will be irrevocably removed upon  
35 the granting of a patent based upon the application.

095416 037.006

Moreover, the designated deposits will be maintained for a period of thirty (30) years from the date of deposit, or for five (5) years after the last request for the deposit; or for the enforceable life of the U.S. patent, whichever is longer. Should a culture become nonviable or be inadvertently destroyed, or, in the case of plasmid-containing strains, lose its plasmid, it will be replaced with a viable culture(s) of the same taxonomic description.

These deposits are provided merely as convenience to those of skill in the art, and are not an admission that a deposit is required under 35 USC §112. The nucleic acid sequences of these plasmids, as well as the amino acid sequences of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the description herein. A license may be required to make, use, or sell the deposited materials, and no such license is hereby granted.

| <u>Strain</u>                               | <u>Deposit Date</u> | <u>ATCC No.</u> |
|---|---------------------|-----------------|
| <i>S. cerevisiae</i> 2150-2-3<br>(pYASI1)   | 2/27/85             | 20745.          |
| <i>S. cerevisiae</i> AB110<br>(pYLUIGF2-14) | 3/19/86             | 20796           |